2018년 2학기 비모수통계자료분석

\# [2016-1] 남녀대학생의 증권시장에 관한 지식의 정도 차이가 존재하는지를 측정하기 위하여 10명씩의 남녀대학생의 표본을 이용하여 25문항의 객관식 기초용어 문제를 출제하여 정답의 점수자료를 아래와 같이 얻었다. 적절한 검정을 행하여라.
(필요시 모든 문제는 유의수준 5% 가정 하에서 해결할 것)

남자대학생 (X)	11	13	22	18	15	10	15	12	19	20
여자대학생 (Y)	10	11	15	18	11	10	16	14	13	17

(1) sign(부호) 검정을 이용하여라. (p값도 계산하여라.)
(2) Median 검정을 이용하여라. (p값도 계산하여라.)

[쌍비교]

X	11	13	22	18	15	10	15	12	19	20
Y	10	11	15	18	11	10	16	14	13	17
	−	−	−		−		+	+	−	−

[가설]

H_0 : $P(+) = P(-)$, 남녀 간 지식의 정도 차이가 존재하지 않는다.
H_1 : $P(+) \neq P(-)$, 남녀 간 지식의 정도 차이가 존재한다.

[검정통계량]
T = "+" 쌍의 총 수 = 2
[유의수준]

$T=2 \sim B(8, \frac{1}{2})$, $P(T \leq 2 | E(T)=np=4)$, $\hat{\alpha} = 0.1445$

[통계적검정]
유의수준 5% 하에서 귀무가설을 채택한다. 따라서 남녀 간 지식의 정도 차이가 존재하지 않는다.

==

[데이터정리]

$\frac{N+1}{2} = \frac{20+1}{2} = 10.5$, $\text{중앙값}(10.5\text{번째값}) = \frac{14+15}{2} = 14.5$

	남	녀	
14.5초과	6(5)	4(5)	10
14.5이하	4(5)	6(5)	10
	10	10	20

[가설]

H_0 : 남녀집단의 중앙값은 동일하다.
H_1 : 남녀집단의 중앙값은 동일하지 않다.

[검정통계량]

$$T = \frac{(6-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(6-5)^2}{5} = 0.8$$

[유의수준]

$\hat{\alpha} = P(T \geq 0.8 | T \sim \chi_1^2)$

$T = 0.8 < \chi^2_{1,0.95} = 3.841$ 이므로 귀무가설을 채택한다.

[통계적검정]
유의수준 5% 하에서 "남녀집단의 중앙값은 동일하다."라는 귀무가설을 기각할 만한 근거가 충분하지 않다. 따라서 남녀집단의 중앙값은 동일하다고 볼 수 있다.

[2016-2] 흡연이 폐암의 직접적인 원인제공의 동기를 믿는지를 알아보기 위하여 대학생 150명을 대상으로 설문조사를 하였다. 또한 흡연의 위험성을 알리는 강의를 수강한 후 흡연과 폐암의 인과관계에 대한 학생들의 의견을 다시 조사한 결과가 아래와 같다. 문항 "흡연은 폐암의 원인인가?"

(1) "강의가 대학생의 의견을 변화시키는 효율적인 방안인가?"를 p값을 이용하여 가설검정 하여라.

강의전 \ 강의후	Yes	No
Yes	30	40
No	60	20

[가설]

H_0 : $P(+) = P(-)$, 강의의 영향력이 없다.
H_1 : $P(+) \neq P(-)$, 강의의 영향력이 있다.

[검정통계량]

T = "+" 쌍의 총 수 = 40

$$T^* = \frac{T - E(T)}{\sqrt{Var(T)}} = \frac{b - np}{\sqrt{npq}} = \frac{b - c}{\sqrt{b+c}} = \frac{40-60}{\sqrt{100}} = \frac{-20}{10} = -2 \sim N(0,1)$$

[유의수준]

$$T = 40 \sim B(100, \frac{1}{2})$$

$$\hat{\alpha} = 2 \times P(T^* \le -2) = 2 \times P(Z \le -2) = 0.046$$

[통계적검정]

유의수준 5% 하에서 "강의의 영향력이 없다."라는 귀무가설을 기각할 만한 근거가 충분하다. 따라서 강의가 대학생의 의견을 변화시키는 효율적인 방안이라고 볼 수 있다.

[2016-3] 전자제품 A, B, C, D 간에 판매액 차이가 존재하는지를 알아보기 위하여 6곳의 대리점을 조사한 결과가(단위 : 천만 원) 다음과 같을 때 전자제품들 간의 판매액 차이가 존재하는지를 밝혀라. Median 검정을 이용하여라. (단 대리점을 블록 취급)

대리점	전자제품			
	A	B	C	D
1	79	87	83	82
2	78	83	81	91
3	84	81	89	80
4	84	76	94	83
5	75	86	90	92
6	81	84	88	85

[데이터정리]

$$\frac{N+1}{2} = \frac{24+1}{2} = 12.5, \quad 중앙값(12.5번째값) = \frac{83+84}{2} = 83.5$$

	A	B	C	D	
83.5 초과	2(2.75)	3(2.75)	4(2.75)	3(2.75)	12
83.5 이하	4(3.25)	3(3.25)	2(3.25)	3(3.25)	12
	6	6	6	6	24

[가설]

H_0 : 전자제품들 간의 판매액 차이가 존재하지 않는다.
H_1 : 전자제품들 간의 판매액 차이가 존재한다.

[검정통계량]

$$T = \frac{(2-3)^2 + (3-3)^2 + (4-3)^2 + (3-3)^2}{3} + \frac{(4-3)^2 + (3-3)^2 + (2-3)^2 + (3-3)^2}{3} = 0.66 + 0.66 = 1.32$$

[유의수준]

$\hat{\alpha} = P(T \ge 1.32 | T \sim \chi^2_3)$, $T = 1.32 < \chi^2_{3,0.95} = 7.815$ 이므로 귀무가설을 채택한다.

[통계적검정]

유의수준 5% 하에서 "전자제품들 간의 판매액 차이가 존재하지 않는다."라는 귀무가설을 기각할 만한 근거가 충분하지 않다. 따라서 전자제품들 간의 판매액 차이가 존재하지 않는다고 볼 수 있다.

[2016-4] 콜레스테롤과 혈압 간에 독립성 여부를 조사하고자 한다. 200명을 랜덤추출하여 얻은 분할자료 결과가 다음과 같다. 두 변수 간의 독립성 검정을 하여라. p값도 구하여라.

	고혈압	정상	저혈압
고콜레스테롤	30	50	20
저콜레스테롤	10	60	30

[데이터정리]

	고혈압	정상	저혈압	
고콜레스테롤	30(20)	50(55)	20(25)	100
저콜레스테롤	10(20)	60(55)	30(25)	100
	40	110	50	200

[가설]

$H_0 : p_{ij} = p_i \times p_j$, 단 $i,j = 1,2$
$H_1 : p_{ij} \neq p_i \times p_j$

[검정통계량]

$$T = \frac{(30-20)^2}{20} + \frac{(50-55)^2}{55} + \frac{(20-25)^2}{25} + \frac{(10-20)^2}{20} + \frac{(60-55)^2}{55} + \frac{(30-25)^2}{25} = 12.91$$

[유의수준]

$\hat{\alpha} = P(T \geq 12.91 \mid T \sim \chi_2^2)$

$T = 12.91 > \chi^2_{2,0.95} = 5.991$ 이므로 귀무가설을 기각한다.

[통계적검정]

유의수준 5% 하에서 귀무가설을 기각할 만한 근거가 충분하다. 따라서 두 변수는 독립이 아니라고 할 수 있다.

[2016-5] 지난해 어느 종합병원의 신생아 탄생은 봄에 55명, 여름에 50명, 가을에 60명, 겨울에 35명이었다. 신생아의 수는 계절에 상관없이 균일한 분포를 이루는지의 chi-square 적합도 검정을 행하여라. p값도 표현하여라.

[자료정리]

	봄	여름	가을	겨울	
O_i	55	50	60	35	$N=200$
E_i	50	50	50	50	

[가설]

H_0 : 신생아의 수는 계절에 상관없이 균일한 분포를 갖는다.

H_1 : 신생아의 수는 계절에 상관없이 균일한 분포를 갖는다.

[검정통계량]

$$T = \frac{(55-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(35-50)^2}{50} = 7$$

[유의수준]

$\hat{\alpha} = P(T \geq 7 \mid T \sim \chi_3^2)$

$T = 7 < \chi^2_{3,0.95} = 7.815$ 이므로 귀무가설을 채택한다.

[통계적검정]

유의수준 5% 하에서 귀무가설을 기각할 만한 근거가 충분하지 않다. 따라서 신생아의 수는 계절에 상관없이 균일한 분포를 갖는다.

[2016-6] 주말에 특정지역의 고속도로를 통과하는 자동차 수를 1분간 10번씩 조사한 자료가 아래와 같다. 자동차 통과 수 X가 정규분포인지의 검정을 수행하여라.

32　33　29　35　33　38　35　35　38　41　단 $\overline{X}=35$, $s=3$을 이용할 것

① Lilliefors 그래프를 이용하여라.

② 부록표를 이용한 검정을 수행하여라.

③ 아래의 분할표를 이용한 적합도 검정을 하여라. 각 cell의 E_i는 25%로 동일함을 가정할 때, cell의 구간임계치가 32.98 / 35 / 37.02 임을 간단히 밝혀라. 추가로 자유도를 신중히 고려할 것.

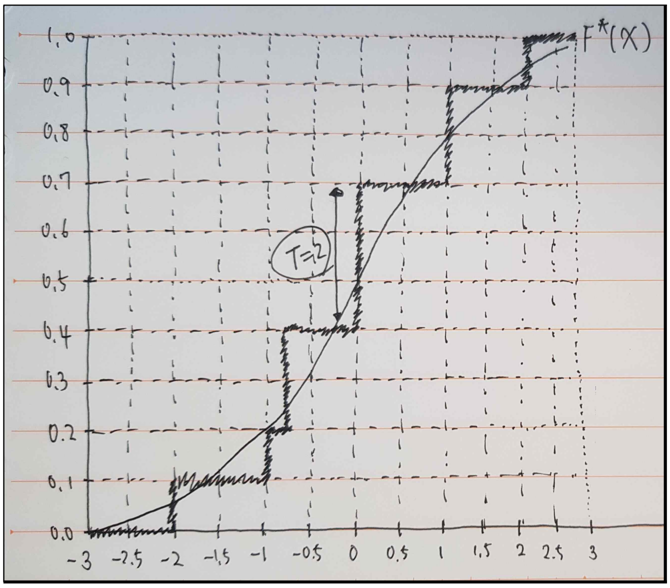
	$X < 32.98$	$32.98 \le X < 35$	$35 \le X < 37.02$	$37.0 \le X$	
O_i					$N=10$
E_i	2.5	2.5	2.5	2.5	

④ 위의 결과가 일치하는지를 밝혀라.

[자료 정렬 및 표준화]

X_i	29	32	33	33	35	35	35	38	38	41
z_i	−2	−1	−0.67	−0.67	0	0	0	1	1	2

[Lilliefors 그래프]



[부록표를 이용한 검정]

$T=0.2 < [\chi^2_{10,0.95}=0.258]$ 이므로 귀무가설을 채택한다. 자동차 통과 수 X가 정규분포를 따른다고 할 수 있다.

[분할표를 이용한 적합도 검정]

	$X < 32.98$	$32.98 \le X < 35$	$35 \le X < 37.02$	$37.0 \le X$	
O_i	2	2	3	3	$N=10$
E_i	2.5	2.5	2.5	2.5	

\# $x_p = \mu + \sigma z_p$

\# $z_{0.25} = -0.6745$ -> $x_{0.25} = 35 - 0.6745(3) = 32.98$

\# $z_{0.50} = 0$ -> $x_{0.50} = 35 - 0(3) = 35$

\# $z_{0.75} = +0.6745$ -> $x_{0.75} = 35 + 0.6745(3) = 37.02$

$T=\dfrac{(2-2.5)^2+(2-2.5)^2+(3-2.5)^2+(3-2.5)^2}{2.5}=0.4$,　$T=0.4<\chi^2_{1,0.95}=3.841$ 이므로 귀무가설을 채택한다.

[결과비교]

두 검정 모두 유의수준 5% 하에서 귀무가설을 채택하는 결과를 나타내었다. 자동차 통과 수 X가 정규분포를 따른다고 말할 수 있다.

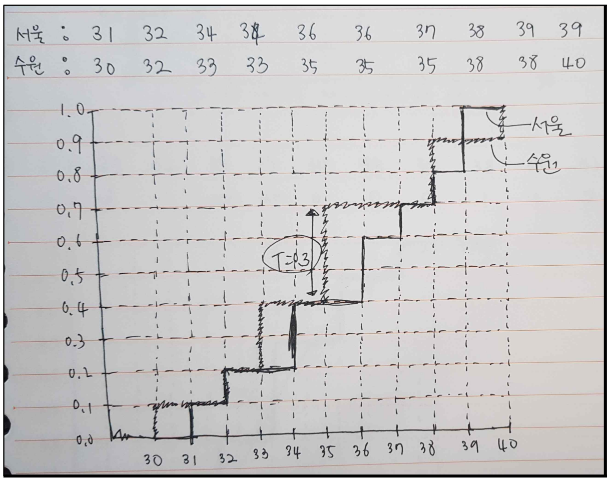
[2016-7] 동일한 전자제품을 취급하는 서울대리점과 수원대리점의 매출액(단위 : 천만 원)이 동일한 분포를 이루는지를 알아보고자 일정기간동안 각각 10번씩의 표본을 추출하여 조사한 자료가 아래와 같다. 두 모집단은 동일한 분포를 갖는지의 Kolmogorov-Smirnov 검정을 수행하여라. 그래프를 이용할 것.

서울대리점	31	32	34	38	34	36	36	37	39	39
수원대리점	32	33	30	35	33	38	35	35	38	40

[가설]

H_0 : 두 대리점의 매출액 분포는 동일분포를 따른다.
H_1 : 두 대리점의 매출액 분포는 동일분포를 따르지 않는다.

[Kolmogorov-Smirnov 검정]



[검정통계량]

$n = m = 10$ 일 때,

유의수준 5% 하에서 $T = 0.3 < [\chi_{0.95} = 0.6]$ 이므로 귀구가설을 채택한다. 즉 두 대리점의 매출액 분포는 동일분포를 따른다.

[2016-8] $n = 4$, $m = 1$ 일 때, “H_0 : 두 모집단의 분포는 동일하다.”는 귀무가설 하에서의 Kolmogorov-Smirnov 이표본 검정통계량 T 의 확률분포표(pdf 및 cdf)를 구하여라. 동시에 $x_{0.05}$ 와 $x_{0.95}$의 분위수값을 구하여라.

시험범위 아님

[2013-1] 주차비용이 인상된 후 자가운전자는 자신의 공공주차장 사용시간(단위 : 분)을 임의로 조사, 기록하였다. 주차시간의 감소추세가 발생하였다고 말할 수 있는가?

8 20 32 28 15 37 33 18 10 15 64 88 22 11 9 17 19 4 10

[데이터정리]

8	20	32	28	15	37	33	18	10	~~15~~
64	88	22	11	9	17	19	4	10	
+	+	−	−	−	−	−	−	.	

[귀무가설]

$H_0:\ P(+) \geq P(-)$, 감소추세가 없다.
$H_1:\ P(+) < P(-)$, 감소추세가 있다.

[검정통계량]

T = "+" 쌍의 총 수 = 2

[유의수준]

$T = 6 \sim B(8, \frac{1}{2})$

$P(T \leq 2 | E(T) = np = 4) = 0.1445$

[통계적검정]

유의수준 5% 하에서 "감소추세가 없다."라는 귀무가설을 기각할 만한 근거가 충분하지 않다. 따라서 주차시간의 감소추세가 발생했다고 보기 어렵다.

[2013-2] 두 대통령 후보 A 및 B 간의 TV 토론이 유권자에게 영향력을 미치는지를 알아보기 위하여 임의로 추출한 69명 유권자의 두 대통령후보에 대한 TV 토론 전 후 선호도의 여론조사를 실시한 자료가 아래와 같았다. McNemar 검정을 실시하여라. p값을 구하여라.

		TV 토론 후 A	TV 토론 후 B	
TV 토론 전	A	17	21	38
	B	15	16	31
		32	37	69

[가설]

$H_0:\ P(+) = P(-)$, TV 토론의 영향력이 없다.
$H_1:\ P(+) \neq P(-)$, TV 토론의 영향력이 있다.

[검정통계량]

T = "+" 쌍의 총 수 = 21

$$T^* = \frac{T - E(T)}{\sqrt{Var(T)}} = \frac{b - np}{\sqrt{npq}} = \frac{b - c}{\sqrt{b + c}} = \frac{21 - 15}{\sqrt{36}} = \frac{6}{6} = 1 \sim N(0,1)$$

[유의수준]

$T = 21 \sim B(36, \frac{1}{2})$

$$\begin{aligned} P(T^* \geq 1) &= 2 \times P(Z \geq 1) \\ &= 2 \times [1 - P(Z \leq 1)] = 2 \times [1 - 0.841] \\ &= 2 \times [0.159] = 0.318 \end{aligned}$$

[통계적검정]

유의수준 5% 하에서 귀무가설을 채택한다. TV 토론의 영향력이 있다고 보기 어렵다.

[2013-3] 현재 사용하는 핸드폰의 고객 만족도 차이유무를 조사하고자 한다. 이를 위하여 3 통신회사 및 고객을 임의로 추출하여 아래의 결과를 얻었다. Median 검정을 행하여라. p값도 구하여라.

통신회사 A	80	79	77	96	88	93	92	80	90
통신회사 B	83	82	91	78	94	71	85	87	
통신회사 C	78	84	89	93	81	75	92	93	

[데이터정리]

$\frac{N+1}{2} = \frac{25+1}{2} = 13$, 중앙값(13번째값) $= 85$

	A	B	C	
85초과	5(4.32)	3(3.84)	4(3.84)	12
85이하	4(4.68)	5(4.16)	4(4.16)	13
	9	8	8	25

[가설]

$H_0:$ 세 모집단의 중앙값은 모두 동일하다.
$H_1:$ 세 모집단의 중앙값은 모두 동일하지 않다.

[검정통계량]

$$T = \frac{(5-4.32)^2}{4.32} + \frac{(3-3.84)^2}{3.84} + \frac{(4-3.84)^2}{3.84} + \frac{(4-4.68)^2}{4.68} + \frac{(5-4.16)^2}{4.16} + \frac{(4-4.16)^2}{4.16} = 0.5720$$

[유의수준]

$\hat{\alpha} = P(T \geq 0.5720 | T \sim \chi_2^2)$

$T = 0.5720 < \chi^2_{2,0.95} = 5.991$ 이므로 귀무가설을 채택한다.

[통계적검정]

유의수준 5% 하에서 "세 모집단의 중앙값은 모두 동일하다."라는 귀무가설을 기각할 만한 근거가 충분하지 않다. 따라서 세 모집단의 중앙값은 모두 동일하다고 볼 수 있다.

[2013-4] 기존의 의약품 A 와 새로 개발한 의약품 B를 50명의 환자에게 임의로 투여한 후 부작용의 반응 여부를 검토하였다. 의약품과 부작용 사이의 독립성 관계가 존재한다는 증거가 있는가?

	부작용(무)	부작용(무)
의약품(A)	30	8
의약품(B)	2	10

[데이터정리]

	부작용(무)	부작용(무)	
의약품(A)	30(24.32)	8(13.68)	38
의약품(B)	2(7.68)	10(4.32)	12
	32	18	50

[가설]

$H_0 : p_{ij} = p_i \times p_j$, 단 $i,j=1,2$

$H_1 : p_{ij} \neq p_i \times p_j$

[검정통계량]

$$T = \frac{(30-24.32)^2}{24.32} + \frac{(8-13.68)^2}{13.68} + \frac{(2-7.68)^2}{7.68} + \frac{(10-4.32)^2}{4.32} = 15.35$$

[유의수준]

$\hat{\alpha} = P(T \geq 15.35 | T \sim \chi_1^2)$

$T = 15.35 < \chi^2_{1,0.95} = 3.841$ 이므로 귀무가설을 채택한다.

[통계적검정]

유의수준 5% 하에서 귀무가설을 기각할 만한 근거가 충분하다. 따라서 의약품과 부작용의 관계는 독립이 아니라고 할 수 있다.

[2013-5] 교육부는 고등학생 신장(키)의 분포가 정규분포인지를 알아보기 위하여 40명의 학생을 임의로 추출하여 키를 측정한 결과가 아래와 같다. 유의수준 5% 하에서 적합도 검정을 하여라.

161	182	141	130	175	188	153	136	178	140
132	133	166	180	163	150	166	133	161	146
124	198	124	142	173	184	138	160	162	188
158	206	133	181	181	190	164	188	140	170

[데이터정리]

	$120 \leq X_i < 142.5$	$142.5 \leq X_i < 165$	$165 \leq X_i < 187.5$	$187.5 \leq X_i < 210$
O_i	12	11	11	6
E_i	10	10	10	10

[통계량 추정]

$\overline{X} = [131.25 \times 12 + 153.75 \times 11 + 176.25 \times 11 + 198.75 \times 6]/40 = 159.93$

$S^2 = [131.25^2 \times 12 + 153.75^2 \times 11 + 176.25^2 \times 11 + 198.75^2 \times 6]/40 - 159.93^2 = 558.95$

$S = \sqrt{558.95} = 23.64$

[가설]

H_0 : 고등학생의 신장의 분포가 정규분포를 따른다.

H_1 : 고등학생의 신장의 분포가 정규분포를 따르지 않는다.

[기댓값]

계급구간값 c_i	$z_p = (c_i - \overline{X})/S$	$F(z_p)$	$E_i = p_i^* N$
120	−1.68	0.05	0.05(40) = 2
142.5	−0.73	0.23	0.18(40) = 7.2
165	0.21	0.58	0.35(40) = 14
187.5	1.16	0.87	0.29(40) = 11.6
210	2.11	0.98	0.11(40) = 4.4
			0.02(40) = 0.8

	$120 \leq X_i < 142.5$	$142.5 \leq X_i < 165$	$165 \leq X_i < 187.5$	$187.5 \leq X_i < 210$
O_i	12	11	11	6
E_i	9.2	14	11.6	5.2

[검정통계량]

$$T = \frac{(12-9.2)^2}{9.2} + \frac{(11-14)^2}{14} + \frac{(11-11.6)^2}{11.6} + \frac{(6-5.2)^2}{5.2} = 1.64$$

$T = 1.64 < \chi^2_{4-1-2,0.95} = 3.841$ 이므로 귀무가설을 채택한다.

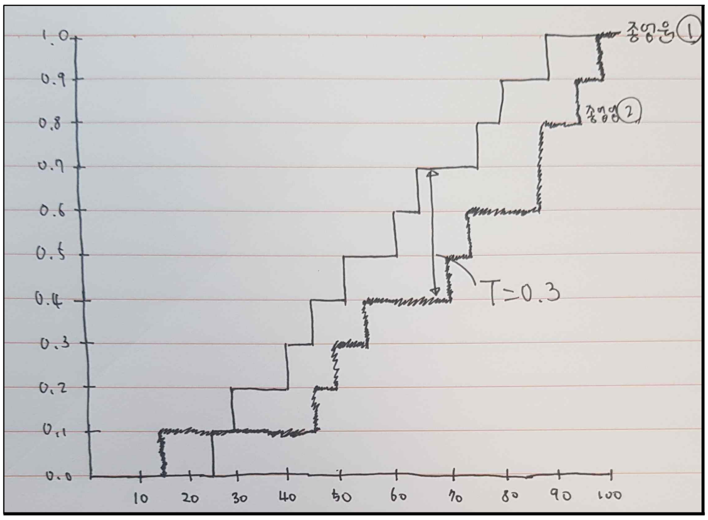
[2013-6] Fast-food 음식점에서 고객의 주문을 완성하는데 걸리는 시간이 두 명의 종업원에 의하여 아래와 같이 각각 오름차순으로 정리하여 기록되었다.(단위 : 초) Kolmogorov 이표본 검정을 수행하여라.

종업원 1	25	30	40	45	52	60	65	75	80	88
종업원 2	15	45	50	55	68	73	85	85	90	95

[가설]

H_0 : 두 종업원의 음식 제조 시간은 동일분포를 따른다.
H_1 : 두 종업원의 음식 제조 시간은 동일분포를 따르지 않는다.

[Kolmogorov 이표본 검정]



[검정통계량 및 통계적검정]

유의수준 5% 하에서 검정통계량 $T=0.3$ 는 임계값 $\chi_{0.95}=0.6$ 보다 작으므로 귀무가설을 채택한다. 따라서 두 종업업의 음식 제조 시간은 동일분포를 따른다.

[2013-7] 다음을 간략히 서술하여라.

① McNemar 검정의 특징은?
② McNemar 검정과 Cox and Stuart 검정의 차이점은?
③ Kolmogorov 일표본 검정이란?
④ Kolmogorov-Smirnov 이표본 검정이란?
⑤ Lillifors의 정규성 검정과정을 간략히 서술하여라.
⑥ 적합도 검정이란?

① McNemar 검정을 위한 자료는 2X2 분할표로 요약되는 이변량의 확률표본 (X_i, Y_i), $i=1,2,\cdots,n^*$로 구성되어 있으며 확률표본 X_i 및 Y_i 는 각각 “0” 및 “1”로 정의할 수 있는 두 범주로 구별되는 특성을 가지고 있다.

② Cox and Stuart 검정은 부호검정의 변횡된 또 다른 응용기법으로 추세의 존재 여부를 파악하는데 용이하다. 따라서 McNemar 검정을 위한 자료가 2X2 분할표라면, Cox and Stuart 검정은 자료들의 수치가 일렬로 나열되어 있는 상태이다.

③ 알려지지 않은 모집단의 누적분포함수 $F(X)$의 가정된 누적분포함수를 $F^*(X)$와 표본추출하여 관측된 자료로부터의 경험적 누적분포함수 $S(X)$를 비교하는 검정기법이다. 검정통계량 T 는 모든 x 값에 대하여 $F^*(X)$와 $S(X)$ 사이의 가장 큰 수직 간의 차이로 구할수 있다.

④ 두 모집단 분포 사이에 존재할 수 있는 위치 및 척도 등의 모든 형태의 차이에 민감한 광범위하여 총괄적인 두 모집단 간의 동일성 여부를 결정하는 검정기법이다.

⑤ 확률표본이 정규분포를 따른다는 귀무가설과 정규분포를 따르지 않는다는 대립가설을 세운 뒤, 검정통계량을 구하기 위해 표본평균 $\overline{X}$ 및 표본표준편차 S 를 구한 뒤 모든 표본관측치를 표준화한다. 그 다음 Lilliefors 그래프를 이용하여 표준화된 관측값의 경험적누적분포함수를 그린다. 표에 제시되지 않은 표본크기에 대해서는 Lilliefors 검정통계량 T 를 표의 가운데 곡선 $F^*(z_i)$와 경험적누적분포함수 $S(z_i)$ 사이의 가장 큰 수직거리로 정의한다. 만약 경험누적분포함수가 Lilliefors 한계선의 어느 한쪽이라도 바깥쪽으로 치우친다면, 모집단은 정규성이 결여되었다고 간주한다. 또한 검정통계량 T 가 표에 제시된 $1-\alpha$ 분위수 보다 크다면 귀무가설을 기각한다.

⑥ 적합도 검정이란 알려지지 않은 분포함수로부터 확률표본을 추출하여 알려지지 않은 분포함수가 사실상 귀무가설로 알려진 분포함수를 따르는지의 적합성 여부를 결정하는 검정기법이다.

[2017-1] 두 종류의 청량음료를 12명의 지원자에게 맛을 보게 한 후, 점수로 맛을 나타내도록 했다. A 청량음료의 맛이 우수한지(높은 점수를 의미)를 유의수준 5% 하에서 부호검정 하여라.

청량음료 A	70	85	73	75	65	50	80	71	80	51	68	77
청량음료 B	65	41	45	80	84	50	71	52	42	78	68	53

[쌍비교]

청량음료 A	70	85	73	75	65	50	80	71	80	51	68	77
청량음료 B	65	41	45	80	84	50	71	52	42	78	68	53
쌍비교	−	−	−	+	+		−	−	−	+		−

[가설]

H_0 : 청량음료 A 가 B보다 맛이 더 우수하지 않다. $P(+) \geq P(-)$

H_1 : 청량음료 A 와 B보다 맛이 더 우수하다. $P(+) < P(-)$

[검정통계량]

T = "+" 쌍의 총 수 = 3

[유의수준]

$T = 3 \sim B(10, \frac{1}{2})$, $P(T \leq 3 | E(T) = np = 5)$, $\hat{\alpha} = P(T \leq 3) = 0.1719$

[통계적검정]

유의수준 5% 하에서 귀무가설을 채택한다. 따라서 청량음료 A 가 B 보다 맛이 더 우수하지 않다.

[2017-2] 흡연이 폐암의 직접적인 원인제공의 동기를 믿는지를 알아보기 위하여 대학생 150명을 대상으로 설문조사를 하였다. 또한 흡연의 위험성을 알리는 강의를 수강한 후 흡연과 폐암의 인과관계에 대한 학생들의 의견을 다시 조사한 결과가 아래와 같다. 문항 "흡연은 폐암의 원인인가?" ① 강의가 대학생의 의견을 변화시키는 방안인가? (유의수준 1% 하에서 양측검정)

강의전 \ 강의후	Yes	No
Yes	33	36
No	64	17

[가설]

H_0 : 강의가 대학생의 의견을 변화시키지 않았다.

H_0 : 강의가 대학생의 의견을 변화시켰다.

[검정통계량]

T = "+" 쌍의 총 수 = 36

$$T^* = \frac{T - E(T)}{\sqrt{Var(T)}} = \frac{b - np}{\sqrt{npq}} = \frac{b - c}{\sqrt{b + c}} = \frac{36 - 64}{\sqrt{100}} = \frac{-28}{10} = -2.8 \sim N(0,1)$$

[유의수준]

$T = 36 \sim B(110, \frac{1}{2})$

$\hat{\alpha} = 2P(T^* \leq -2.67) = 2P(Z \leq -2.67) = 0.006$

[통계적검정]

유의수준 5% 하에서 "강의가 대학생의 의견을 변화시키지 않았다."라는 귀무가설을 기각할 만한 근거가 충분하다. 따라서 강의가 대학생의 의견을 변화시키는 효율적인 방안이라고 볼 수 있다.

[2017-3] 남녀대학생의 현재 사용하는 동일 핸드폰의 고객 만족도 차이유무를 조사하기 위하여 10명씩의 남녀대학생의 표본을 이용하여 100점 만점의 만족도 점수자료를 아래와 같이 얻었다. 유의수준 5% 하에서 적절한 검정을 행하여라. Median 검정을 이용하라.

여자대학생(X)	89	82	91	78	94	71	97	86	87	78
남자대학생(Y)	78	84	83	93	81	75	97	92	97	80

[분할표]

$\frac{N+1}{2} = \frac{20+1}{2} = 10.5$, $\text{중앙값}(10.5\text{번째 값}) = \frac{84+86}{2} = 85$

	여자	남자	
85초과	6(5)	4(5)	10
85이하	4(5)	6(5)	10
	10	10	20

[가설]

H_0 : 남녀대학생 간의 중앙값이 차이가 없다.
H_1 : 남녀대학생 간의 중앙값이 차이가 있다.

[검정통계량]

$$T = \frac{(6-5)^2 + (4-5)^2 + (4-5)^2 + (6-5)^2}{5} = \frac{4}{5} = 0.8$$

[통계적검정]

유의수준 5% 하에서 검정통계량 $T = 0.8 < \chi^2_{1,0.95} = 3.841$ 이므로 귀무가설을 채택한다. 남학생과 여학생에 따른 과목의 선호도의 차이가 없다.

[2017-4] 남학생과 여학생간에 국어, 수학, 영어 세과목에 대한 선호도가 다른가를 조사하고자 한다. 남학생 250명과 여학생 250명을 랜덤 추출하여 가장 좋아하는 한 과목을 선택하게 하여 분류한 결과가 다음과 같다. 남학생과 여학생에 따른 과목의 선호도가 다르다고 할 수 있는가?

	국어	수학	영어
남학생	70	80	100
여학생	80	60	110

[가설]

H_0 : 남학생과 여학생에 따른 과목의 선호도의 차이가 없다.
H_1 : 남학생과 여학생에 따른 과목의 선호도의 차이가 있다.

[데이터정리]

	국어	수학	영어	
남학생	70(75)	80(70)	100(105)	250
여학생	80(75)	60(70)	110(105)	250
	150	140	210	500

[검정통계량]

$$T = \frac{(70-75)^2 + (80-75)^2}{75} + \frac{(80-70)^2 + (60-70)^2}{70} + \frac{(100-105)^2 + (110-105)^2}{105} = 4$$

$T = 4 \sim \chi^2_2$

[통계적검정]

$T = 4 < \chi^2_{2,0.95} = 5.991$ 이므로 귀무가설을 채택한다. 따라서 남학생과 여학생에 따른 과목의 선호도의 차이가 없다고 할 수 있다.